

### **REMARKS**

In the non-final Office Action, the Examiner rejects claims 1-22, 30, and 31 under 35 U.S.C. § 101; rejects claims 1, 4-11, 14-18, 21, 22, and 30 under 35 U.S.C. § 102(b) based on the non-patent publication “Fast Speaker Change Detection for Broadcast News Transcription and Indexing”, Daben Liu et al. (referred to as “Liu” herein); rejects claims 2, 12, and 19 under 35 U.S.C. § 103(a) based on Liu in view of the non-patent publication “A Distance Measure Between Collections of Distributions and its Application to Speaker Recognition,” Homayoon Beigi et al. (referred to as “Beigi” herein); rejects claims 3, 13, 20, and 21 under 35 U.S.C. § 103(a) based on Liu in view of U.S. Patent No. 6,317,716 to Braida et al. (“Braida”); rejects claims 23-25 and 27-29 under 35 U.S.C. § 103(a) based on Liu in view of the non-patent publication “Spoken Documents: Creating Searchable Archives from Continuous Audio,” Sean Colbath et al. (referred to as “Colbath” herein”) and further in view of Braida; and rejects claim 26 under 35 U.S.C. § 103(a) based on Liu, Colbath, and Braida, and further in view of Beigi. These rejections are respectfully traversed.

By this Amendment, Applicants amend claims 1, 4, 11, 18, 23, and 30 to improve form.

#### *Rejections under 35 U.S.C. § 101*

Claims 1-22, 30, and 31 stand rejected under 35 U.S.C. § 101 as being, according to the Examiner, directed to non-statutory subject matter. According to the Examiner, these claims define “merely a series of steps ” without any “claimed limitation to a practical application.” Applicants do not agree with the Examiner’s assertion. Claim 1, for example, is directed to a method that includes, among other things, “detecting speaker changes.” As described in the instant application,

at, for example, paragraphs 0005 and 0006, detecting speaker changes in an audio stream is a desirable and a practical application.

In any event, without acquiescing in the Examiner's rejection, but in order to expedite prosecution, Applicants have amended independent claims 1, 11, 18, and 30 to recite a clearly tangible result. These claims therefore include a "claimed limitation to a practical result." Claim 1, for example, now recites "outputting an indication of the detected speaker changes," claim 11 recites "store an indication of the detected speaker changes," claim 18 recites that "an indication of the detected locations of speaker changes are output from the device," and claim 30 recites "means for outputting the detected speaker changes." Applicants submit that the rejection of these claims under 35 U.S.C. § 101 is thus clearly improper and should be withdrawn.

*Rejections under 35 U.S.C. § 102(b) Based on Liu*

Claims 1, 4-10, 11, 14-17, 18, 21, 22, and 30 stand rejected under 35 U.S.C. § 102(b) based on Liu. For the following reasons, Applicants respectfully traverse this rejection.

Amended claim 1 is directed to a method for detecting speaker changes in an input audio stream. The method includes segmenting the input audio stream into predetermined length intervals; decoding the intervals to produce a set of phones corresponding to each of the intervals; generating a similarity measurement based on a first portion of the audio stream that is within one of the intervals and that occurs prior to a boundary between adjacent phones in one of the intervals and a second portion of the audio stream that is within the one of the intervals and that occurs after the boundary; detecting speaker changes based on the similarity measurement; and outputting an indication of the detected speaker changes.

Liu does not disclose or suggest each of the features recited in claim 1. Liu, for example, does not segment the input audio stream into predetermined length intervals, as recited in claim 1.

The Examiner contends that Liu discloses this feature of claim 1 and particularly points to the first paragraph of section 4 (Speaker Change Detection) of Liu. (Office Action, page 3.) This paragraph of Liu describes speaker change detection as it was implemented in a previous system, called the “BBN Byblos system.” The remaining paragraphs in section 4, of Liu, however, describe the speaker change detection algorithm that is the subject of the Liu publication. The speaker change detection algorithm that is the subject of the Liu publication, however, does not disclose or suggest segmenting an input audio stream into predetermined length intervals. In contrast, as shown in the flow chart of Fig. 2 of Liu and as described in the corresponding description, Liu discloses using a variable size window that is generated by simply accumulating phonemes until two seconds of audio is buffered. That is, in Liu, the variable size window is incremented “one phone at a time” and then used to search for speaker changes on phone boundaries in the window. The variable size window that is the subject of the Liu publication does not disclose or suggest segmenting the input audio stream into predetermined length intervals, as recited in claim 1.

Claim 1 further recites generating a similarity measurement based on a first portion of the audio stream that is within one of the intervals and that occurs prior to a boundary between adjacent phones in one of the intervals and a second portion of the audio stream that is within the one of the intervals and that occurs after the boundary. Liu also does not disclose or suggest this feature of claim 1.

The Examiner contends that section 4 of Liu discloses this feature of claim 1. (Office Action, page 3.) Applicants disagree with the Examiner’s interpretation of Liu.

As previously mentioned, section 4 of Liu, and in particular the speaker change detection flow chart shown in Fig. 2 of Liu, detects speaker changes within a variable size window. Within each variable size window, Liu calculates values  $\lambda'$  and finds the position in each window where  $\lambda'$  is a maximum. Liu, however, does not disclose or suggest, as recited in claim 1, generating a similarity measurement based on a first portion of the audio stream that is within one of the intervals and that occurs prior to a boundary between adjacent phones in one of the intervals and a second portion of the audio stream that is within the one of the intervals and that occurs after the boundary.

For at least these reasons, Applicants submit that Liu does not disclose or suggest each of the features recited in claim 1. Therefore, the rejection of claim 1 under 35 U.S.C. § 102(b) based on Liu is improper and should be withdrawn. The rejection of claims 4-10, at least by virtue of their dependency from claim 1, is also improper and should be withdrawn.

Independent claim 11 and its dependent claims 14-17 also stand rejected under 35 U.S.C. § 102(b) based on Liu.

Amended claim 11 is directed to a device for detecting speaker changes in an audio signal. The device includes a processor and a memory. The memory contains instructions that when executed by the processor cause the processor to segment the audio signal into predetermined length intervals; decode the intervals to produce a set of phones corresponding to each of the intervals, generate a similarity measurement based on a first portion of the audio signal that occurs prior to a boundary between phones in one of the sets of phones of an interval and a second portion of the audio signal that occurs after the boundary; detect speaker changes based on the similarity measurement; and store an indication of the detected speaker changes.

Liu does not disclose or suggest each of the features recited in claim 11. Liu, for example, does not segment an audio signal into predetermined length intervals. As discussed previously with respect to claim 1, the speaker change detection algorithm that is the subject of the Liu publication does not disclose or suggest segmenting an audio signal into predetermined length intervals. In fact, as shown in the flow chart of Fig. 2 of Liu and as described in the corresponding description, Liu discloses using a variable size window that is incremented “one phone at a time” to search for speaker changes on each phone boundary.

Claim 11 further recites generating a similarity measurement based on a first portion of the audio signal that occurs prior to a boundary between phones in one of the sets of phones of an interval and a second portion of the audio signal that occurs after the boundary. Liu also does not disclose or suggest this feature of claim 11. Section 4 of Liu, for instance, and in particular the speaker change detection flow chart shown in Fig. 2 of Liu, relates to detecting speaker changes within a variable size window. Within each variable size window, Liu calculates values  $\lambda'$  and finds the position in each window where  $\lambda'$  is a maximum. Neither this section of Liu nor any other section of Liu, however, discloses or suggests generating a similarity measurement based on a first portion of the audio signal that occurs prior to a boundary between phones in one of the sets of phones of an interval and a second portion of the audio signal that occurs after the boundary.

For at least these reasons, Applicants submit that Liu does not disclose or suggest each of the features recited in claim 11. Therefore, the rejection of claim 11 under 35 U.S.C. § 102(b) based on Liu is improper and should be withdrawn. The rejection of claims 14-17, at least by virtue of their dependency from claim 11, is also improper and should be withdrawn.

Independent claim 18 and its dependent claims 21 and 22 also stand rejected under 35 U.S.C. § 102(b) based on Liu. Claim 18 includes certain features similar to, although possibly of different scope than, those recited in claims 1 and 11. Accordingly, for reasons similar to those given above for claims 1 and 11, Applicants submit that the rejection of claim 18 under 35 U.S.C. § 102(b) based on Liu is improper and should be withdrawn. The rejection of claims 21 and 22, at least by virtue of their dependency from claim 18, is also improper and should be withdrawn.

Independent claim 30 also stands rejected under 35 U.S.C. § 102(b) based on Liu. Claim 30 includes certain features similar to, although possibly of different scope than, those recited in claim 1. Accordingly, for reasons similar to those given above for claim 1, Applicants submit that the rejection of claim 30 under 35 U.S.C. § 102(b) based on Liu is improper and should be withdrawn.

*Rejection under 35 U.S.C. § 103(a) Based on Liu and Beigi*

Claims 2, 12, and 19 stand rejected under 35 U.S.C. § 103(a) based on Liu and Beigi. For the following reasons, Applicants respectfully traverse this rejection.

Applicants have reviewed Beigi and submit that Beigi does not cure the previously mentioned deficiencies of Liu. Therefore, at least by virtue of the dependency of these claims from 1, 11, and 18, respectively, Applicants submit that the rejection of these claims are improper and should be withdrawn.

*Rejection under 35 U.S.C. § 103(a) Based on Liu and Braida*

Claims 3, 13, 20, and 31 stand rejected under 35 U.S.C. § 103(a) based on Liu and Braida. For the following reasons, Applicants respectfully traverse this rejection.

Claim 3 depends from claim 1 and recites creating the predetermined length intervals such that portions of the intervals overlap one another. The Examiner relies on Braida, and specifically cites column 7, lines 7-15 of Braida, to disclose this feature of claim 3. (Office Action, page 7). Applicants respectfully disagree with this rejection.

The cited section of Braida describes a “parameterization program 40” that retrieves speech samples in frames of “200 samples with 100-sample overlap between successive frames.” Applicants submit that neither this section of Braida nor any other section of Braida suggests modifying Liu to include the “sample overlap” described by Braida. The technique of Liu describes using a variable size window in which detected speaker boundary points are used as breaks between windows. (See Liu, section 4, “Speaker Change Detection”). If anything, Liu explicitly discloses non-overlapping windows and thus teaches away from the modification suggested by the Examiner. Accordingly, one of ordinary skill in the art reading Liu and Braida would not be motivated to combine Liu and Braida in the manner suggested by the Examiner.

For at least this additional reason, Applicants submit that the Examiner has not made a *prima facie* case of obviousness with respect to claim 3. Accordingly, the rejection of claim 3 based on Liu and Braida is improper and should be withdrawn.

Dependent claims 13, 20, and 31 recite features similar to those recited in claim 3. For reason similar to those given for claim 3, Applicants submit that the rejection of these claims based on Liu and Braida is also improper and should be withdrawn.

*Rejection under 35 U.S.C. § 103(a) Based on Liu, Colbath and Braida*

Claims 23-25 and 27-29 stand rejected under 35 U.S.C. § 103(a) based on Liu, Colbath, and Braida. For the following reasons, Applicants respectfully traverse this rejection.

Amended claim 23 is directed to a system comprising an indexer configured to receive input audio data and generate a rich transcription from the audio data, the rich transcription including metadata that defines speaker changes in the audio data. The indexer includes a segmentation component configured to divide the audio data into overlapping segments of a predetermined length, speaker change detection component configured to detect locations of speaker changes in the audio data based on a similarity value calculated at locations in the segments that correspond to phone class boundaries; a memory system for storing the rich transcription; and a server configured to receive requests for documents and to respond to the requests by transmitting ones of the rich transcriptions that match the requests. Liu, Colbath, and Braida, either alone or in combination, do not disclose or suggest the features of this claim.

Neither Colbath nor Braida suggest the modification of Liu to include, for example, a segmentation component configured to divide the audio data into overlapping segments of a predetermined length. As discussed with respect to claim 1, the speaker change detection algorithm that is the subject of the Liu publication does not disclose or dividing the audio data into overlapping segments of a predetermined length. In fact, as shown in the flow chart of Fig. 2 of Liu and as described in the corresponding description, Liu discloses using a variable size window that is incremented “one phone at a time” to search for speaker changes on each phone boundary.



Further, regarding the segmentation component recited in claim 23, the Examiner additionally relies on Braida to disclose a “system that uses overlapping frames.” (Office Action, page 8, citing column 7, lines 7-15 of Braida.)

Column 7, lines 7-15 of Braida describe a “parameterization program 40” that retrieves speech samples in frames of “200 samples with 100-sample overlap between successive frames.” Applicants submit that neither this section of Braida nor any other section of Braida suggests modifying Liu to include the “sample overlap” described by Braida. The technique of Liu describes using a variable size window in which detected speaker boundary points are used as breaks between windows. (See Liu, section 4, “Speaker Change Detection”). Thus, if anything, Liu’s explicit disclosure of non-overlapping windows teaches away from the modification suggested by the Examiner. Accordingly, one of ordinary skill in the art reading Liu and Braida would not be motivated to combine Liu and Braida in the manner suggested by the Examiner.

For at least these reasons, Applicants submit that Liu and Braida, either alone or in combination, do not disclose or suggest the segmentation component recited in claim 23. Applicants submit that Colbath does not cure this deficiency of Liu and Braida.

For at least these reasons, Applicants submit that the rejection of claim 23 based on Liu, Colbath, and Braida is improper and should be withdrawn. The rejection of claims 24, 25, and 27-29, at least by virtue of their dependency from claim 23, are also improper and should be withdrawn.

*Rejection under 35 U.S.C. § 103(a) Based  
on Liu, Colbath, Braida and Beigi*

Claim 26 stands rejected under 35 U.S.C. § 103(a) based on Liu, Colbath, Braida, and Beigi.

Applicants have reviewed Beigi and submit that Beigi does not cure the previously mentioned deficiencies of Liu, Colbath, and Braida. Therefore, at least by virtue of the dependency of claim 26 from claim 25, Applicants submit that the rejection of claim 26 is improper and should be withdrawn.

**CONCLUSION**

In view of the foregoing amendments and remarks, Applicants respectfully request the Examiner's reconsideration of the application and the timely allowance of the pending claims.

As Applicants' remarks with respect to the Examiner's rejections are sufficient to overcome these rejections, Applicants' silence as to certain assertions by the Examiner in the Office Action or certain requirements that may be applicable to such rejections (e.g., whether a reference constitutes prior art, motivation to combine references, assertions regarding dependent claims, etc.) is not a concession by Applicants that such assertions are accurate or such requirements have been met, and Applicants reserve the right to analyze and dispute these assertions/requirements in the future.

Applicant believes no fee is due with this response. However, if a fee is due, please charge our Deposit Account No. 18-1945, under Order No. BBNT-P01-086 from which the undersigned is authorized to draw.

Dated: July 2, 2007

Respectfully submitted,

By 

Edward A. Gordon

Registration No.: 54,130

FISH & NEAVE IP GROUP, ROPES & GRAY  
LLP

One International Place

Boston, Massachusetts 02110

(617) 951-7000

(617) 951-7050 (Fax)

Attorneys/Agents For Applicant